The global magnitude-frequency relationship for large explosive volcanic eruptions

Jonathan Rougier[a] R. Stephen J. Sparks[b]

Katharine V. Cashman[b] Sarah K. Brown[b]

Abstract

For volcanoes, as for other natural hazards, the frequency of large events diminishes with their magnitude, as captured by the magnitude-frequency relationship. Assessing this relationship is valuable both for the insights it provides about volcanism, and for the practical challenge of risk management. We derive a global magnitude-frequency relationship for explosive volcanic eruptions of at least 300 Mt of erupted mass (or M4.5). Our approach is essentially empirical, based on the eruptions recorded in the LaMEVE database. It differs from previous approaches mainly in our conservative treatment of magnitude-rounding and under-recording. Our estimate for the return period of 'super-eruptions' (1000 Gt, or M8) is 17 ka (95% CI: 5.2 ka, 48 ka), which is substantially shorter than previous estimates, indicating that volcanoes pose a larger risk to human civilisation than previously thought.

Keywords: geohazard, extreme event, LaMEVE, exceedance probability, return period, marked Poisson process

[a]School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK. Email `j.c.rougier@bristol.ac.uk`. Corresponding author.

[b]School of Earth Sciences, University of Bristol, Wills Memorial Building, Queens Road, Bristol BS8 1RJ, UK. Emails: `steve.sparks@bristol.ac.uk`, `kathy.cashman@bristol.ac.uk`, and `sarah.k.brown@bristol.ac.uk`.

1 Introduction

There are both fundamental science reasons and practical reasons for establishing a global relationship between magnitude and frequency for explosive volcanic eruptions. The magnitude-frequency relationship constrains rates of volcanism, provides potential insights into the underlying tectonic and igneous processes that control volcanism and establish the conditions for explosive eruptions, and provides critical information to forecast future eruptions and assess attendant volcanic hazards, including the effects on climate of large explosive eruptions.

More broadly, interest in extreme geohazard events and their consequences is increasing following a series of high-profile earthquakes, tropical cyclones and tsunamis that have had substantial regional impacts (e.g., Plag et al., 2015). From this perspective, the frequency of very large explosive eruptions is of particular importance due to the potential for such eruptions to have not only regional but also global environmental and societal effects. Although the magnitude-frequency relationship for large-magnitude eruptions has been well-studied (Pyle, 1995; Siebert et al., 2010; Deligne et al., 2010; Sheldrake and Caricchi, 2017), some uncertainty remains, while the relationship for the largest-magnitude explosive eruptions is not well known (although see Mason et al., 2004).

The challenge for estimating the magnitude-frequency relationship is that large explosive eruptions are rare. Records of the largest eruptions are extracted from proxies in geological archives. Naturally, such proxies are are hard to interpret, and the resulting values for dating and magnitude have substantial uncertainties and may be systematically biased. The frequency of eruptions in a modern database is also misleading, because the probability of an historical eruption leaving a trace that survives to be found and included in the database depends on the time, location, and magnitude of the eruption. Thus, incautious use of recorded large eruptions can lead to an inaccurate estimate of the magnitude-frequency relationship. Our approach in this paper is conservative with respect to mis-

recording, and all of our point estimates are accompanied by 95% confidence or credible intervals.

The plan of the paper is as follows. Section 2 describes the scale for magnitude, and two complementary ways to present the magnitude-frequency relationship: the exceedance probability curve and the return period curve. Section 3 describes the database and the records it contains, highlighting two sources of inaccuracy. Section 4 describes our statistical model, and uses it to estimate a semi-parametric approximation of the exceedance probability curve. Section 5 introduces a parametric model better able to accommmodate the limitations in the records. Section 6 presents our preferred estimate of the exceedance probability curve, based on the parametric model, and compares our estimates of the return period with others in the literature. Section 7 concludes with a summary and a brief discussion of the implications of our estimate.

2 The magnitude-frequency relationship

The magnitude scale is

$$M = \log_{10}(\text{erupted mass in kg}) - 7, \tag{1}$$

as defined by Pyle (2000) and Mason et al. (2004). We prefer this scale to the widely used Volcanic Explosivity Index (VEI, see Newhall and Self, 1982) because VEI is ordinal and so cannot be represented by a continuous function to describe magnitude and frequency. Further, VEI is assigned to an eruption based on multiple criteria, including eruption column height, which cannot be directly related to magnitude, so VEI is not consistently a measure of magnitude. However, the legacy of VEI creates difficulties in interpreting records of previous eruptions, as discussed in section 3.

The global magnitude-frequency relationship for large explosive eruptions can

be represented in two complementary ways. First, in terms of the 'exceedance probability' curve, here denoted $\bar{P}$. The value $\bar{P}(m)$ is the probability of at least one eruption of at least magnitude m happening somewhere in the world in the next year. The largest recorded eruption since 100 ky is Toba (Indonesia), dated 73 ky, recorded at $M = 9.1$ (Costa et al., 2014). The value $\bar{P}(9.1)$ is the probability of another Toba (or worse) happening in the next year. In this paper we use 'My' and 'ky' to denote a point in time in years BP, and 'Ma' and 'ka' to denote a duration.

Second, the magnitude-frequency relationship can be represented in terms of the 'return period' curve, denoted R. The value $R(m)$ is the mathematical expectation of the time to wait until an eruption with magnitude of at least m. Thus $R(9.1)$ is the expected time to wait, in years, until an eruption which is at least as large as Toba.

Both the exceedance probability curve and the return period curve can be derived within a stochastic process model for eruption times and magnitudes. In our marked Poisson process model they are complementary, because $R(m) \approx 1/\bar{P}(m)$ if $\bar{P}(m)$ is small (see section 6). However, the two labels '$\bar{P}(m) = 0.001$' and '$R(m) = 1000$ years' will often be interpreted differently by non-experts. The latter seems more user-friendly, but can give a very misleading impression, particularly in a changing environment (although this is more relevant to flooding than to volcanoes).

There is another reason for preferring exceedance probabilities over return periods, which is both technical and practical. The time to wait until an eruption is an unbounded quantity, and consequently the value of its expectation is susceptible to very large values occurring with small probabilities; in fact, the expectation may be infinite, particularly when integrating out the parameters in a Bayesian approach. This is a general problem with expectations: they can provide poor summary values for unbounded quantities. Therefore, we prefer to represent the magnitude-frequency relationship as the exceedance probability curve. Where re-

turn periods are required, we adopt the convention of using the reciprocal of the exceedance probability, providing that this probability is small.

3 The volcanic record

The Large Magnitude Explosive Volcanic Eruptions database (LaMEVE) provides a global compilation of data on magnitudes and ages during the Quaternary (Crosweller et al., 2012; Brown et al., 2014). LaMEVE has been developed to complement the Volcanoes of the World (VOTW) database of the Smithsonian Institution for the Holocene and is based on literature for pre-Holocene entries. This analysis is based on version 3.1 of the database, relased in Oct. 2015. However, in the light of our preliminary results we initiated a revision of all records of eruptions since 100 ky with $M \geq 7$, and some uncertain records at lower magnitudes. The results will be incorporated into the next version of LaMEVE, but in the meantime our dataset is available as a spreadsheet in the supplementary information to this paper.

This paper focuses on records in LaMEVE that are dated to have occurred since 100 ky, 1379 eruptions in total. This section considers the difficulties in interpreting these records. One difficulty which we need not consider, except in passing, is the challenge of dating an eruption from its trace in the geological record. This is because we sidestep dating uncertainty by using a statistical model which is time-invariant, at the global scale. This 'stationarity' assumption is discussed in more detail in section 4.

3.1 Magnitude accuracy

Pyle (2016) summarises the methods for assessing magnitude from geological data, and the many sources of error, and thus of uncertainty. He does not provide uncertainty estimates. However, an assessment of volume estimates from isopach maps of tephra fall deposits with at least 20 data thickness points indicates uncertainties

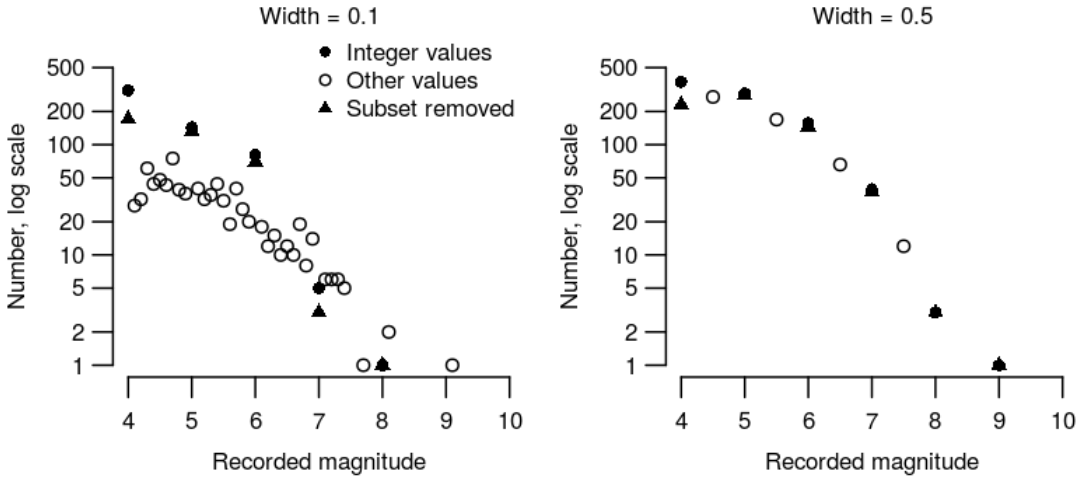


Figure 1: Recorded magnitudes from the LaMEVE database, for eruptions dated to have occurred since 100 ky, using the magnitude scale of Pyle (2000), expressed to the nearest 0.1. The vertical scale is logarithmic. The lefthand panel show the raw values; the righthand panel shows the values in bins of width 0.5. There is strong evidence of rounding to the nearest integer, even after removing a subset of values for which rounding is known to be present.

typically exceeding $M \pm 0.3$ (Engwell et al., 2015).

Measurement errors are fairly unsystematic, being a source more of noise than of bias. However, inspection of the frequencies of recorded magnitudes reveals a systematic error and thus a potentially large source of bias. The lefthand panel of Figure 1 shows that recorded frequencies pile-up on the integer magnitude values, which must be an artefact; see also Brown et al. (2014).

By going back through the database and the supporting papers, we identified one source of rounding. A subset of the records are eruptions with a recorded VEI of v (an integer) but without a reported magnitude, and these were coded as $M = v.0$. However, a VEI value of v corresponds to a magnitude of $v.0$ to $v.9$. There were 163 such eruptions in records dated since 100 ky. This is 'rounding down', which shifts the exceedance probability downwards, understating the exceedance probability of large explosive eruptions, and overstating the length of the return period for large explosive eruptions.

Figure 1 also shows the frequencies of recorded magnitudes after removing the subset identified above. The frequencies still pile-up on the integer magnitude

values, indicating that there is another source of rounding. The righthand panel of Figure 1 shows that widening the bins from width 0.1 to width 0.5 does not remove the piling up. We suspect that this source is rounding towards the nearest integer. We speculate—and it is no more than that—that a volcanologist who assesses a magnitude that is close to an integer may well round to the integer, in the light of her own assessment of uncertainty, in order not to give a spurious impression of accuracy. However, as a reviewer notes, there is an issue about whether the volcanologist assesses volume and then rounds, and then the rounded value is converted to mass using a standard density such as 2500 kg/m^3, or whether the volcanologist assesses mass directly and rounds that. In due course a better operational understanding of rounding might change our results. We return to this topic in the discussion of Table 2 in section 6.

In order to make progress, we will group the recorded magnitudes into integer-width bins centred at the integers, reflecting our view, supported by Figure 1, that rounding to the nearest integer is the dominant source of piling-up on the integer magnitude values. Any aggregation into bins will reduce the effect of rounding, even if it does not remove it completely. We will exclude recorded magnitudes below $M = 4.5$ for which there is no integer-width bin, because the LaMEVE database is for $M \geq 4$. Further screening for under-recording, described immediately below, removes all but one of the records in the rounding-down subset identified above, so that they no longer contribute downward bias to the exceedance probability. The one remaining record from this subset is a VEI = 6 eruption from an unknown source, dated 1808 CE, which we recoded as $M = 6.3$.

3.2 Under-recording

The second source of error is variations in the recording probability, which is the probability that a past eruption appears in the LaMEVE database. Figure 2 shows a simple diagnostic of under-recording by magnitude and time (see, e.g.,

Guttorp and Thompson, 1991; Rougier et al., 2016). Under the hypothesis that the eruption rate in a magnitude bin is effectively time-invariant, non-linearity in the cumulative number of eruptions through time indicates that the recording probability varies in time, and convexity indicates that it decreases going back in time.

Figure 2 shows that under-recording is a serious problem in the database, and that the recording probability varies by magnitude, and—broadly speaking—decreases going back in time, as would be expected. The scale and nature of the under-recording casts doubt on studies extending back over the last 100 ka which have made no adjustment for under-recording, particularly those which claim to find differences in eruption behaviour by magnitude, when this could easily reflect differences in under-recording by magnitude (see, e.g., Tatsumi and Suzuki-Kamata, 2014).

For the bin $7.5 \leq M$, the gaps in the recorded eruptions are suggestive of unrecorded eruptions. The compelling evidence of substantial under-recording at lower magnitudes makes this a simpler explanation than invoking some kind of episodic tectonic process. Below, we will allow for the possibility of missing eruptions. This is an advantage of using a time-invariant binned approach: it is easy to adjust for specified instances of possible under-recording without having to consider when the missing eruptions occurred, and precisely how large they were.

Now consider the smaller-magnitude bins in Figure 2. The recording probability is currently 1, for all large magnitudes, in our populous era of global monitoring. Thus the first upward bend, elbow, or gap, going back in time from now, suggests the time at which the recording probability drops substantially below 1. Figure 3 zooms-in to the recent past, and identifies, by eye, the point at which the recording probability can be taken to be effectively 1, for $M < 7.5$. Deciding on the precise timing of an abstract event is always going to be subjective, but we believe that the human eye, aided by our knowledge of recording practices,

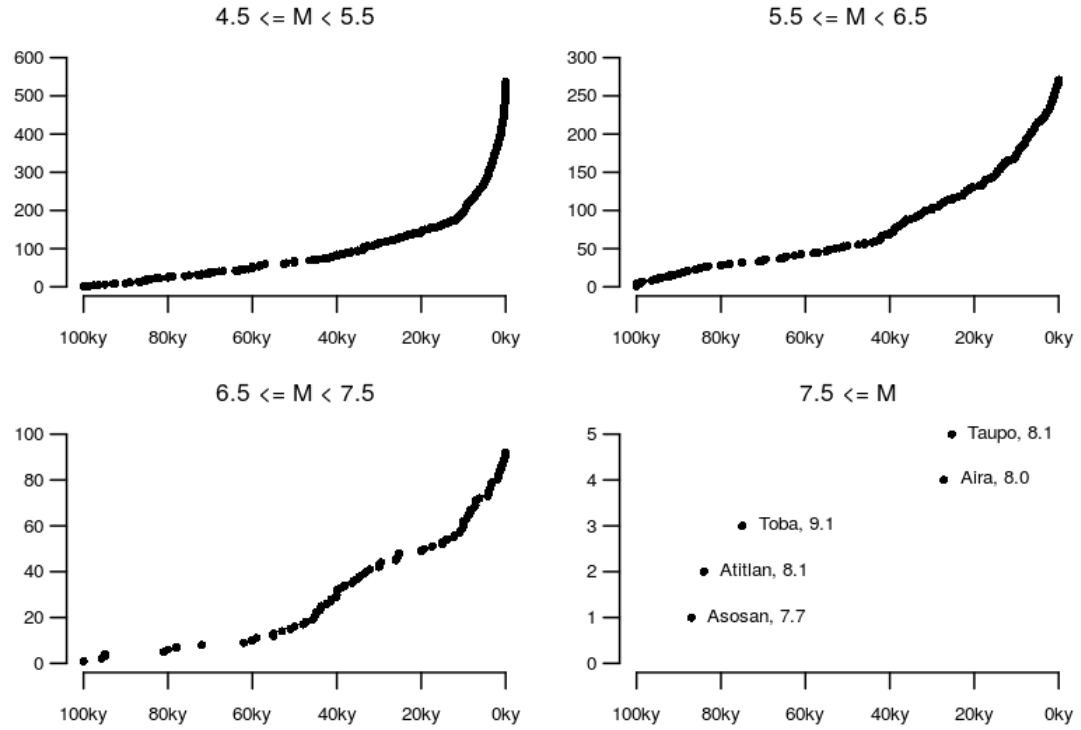


Figure 2: Cumulative numbers of eruptions by time, where each panel shows a different magnitude bin. Convexity indicates that the recording probability decreases going back in time. In the final panel, the eruptions are shown with their names and magnitudes.

is more refined than a statistical test. In any event, the precise location of the vertical lines is not important, because a recording probability of a little below 1 is close enough, given our inability to make fine distinctions about the eruption rates.

205 This screening for under-recording drops a large number of records (the numbers remaining are given in Figure 4). In our analysis we favour reducing bias and carefully quantifying variability, because the alternative, a downward-biased estimate of the exceedance probability curve with small variability, could be seriously misleading.

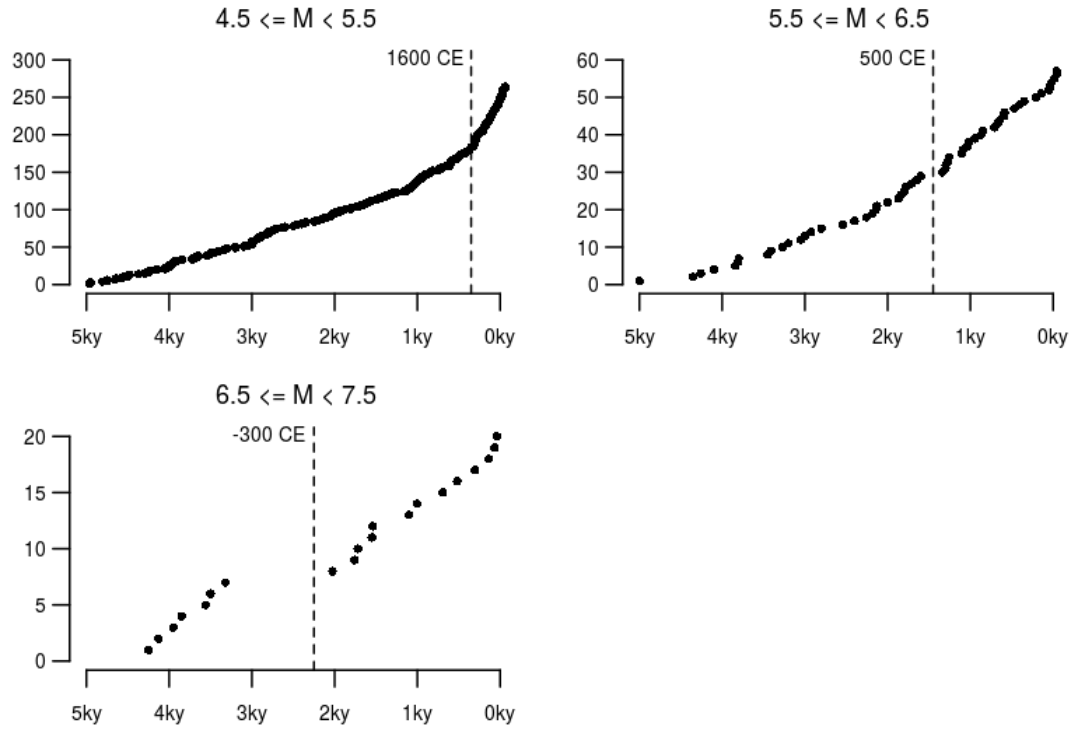


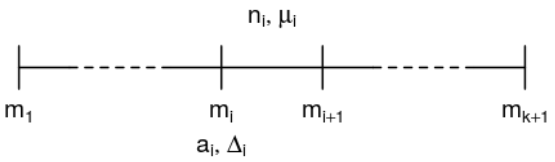
Figure 3: The same as Figure 2, except zoomed in to the most recent 5 ka. The vertical dashed line indicates the point in time at which the recording probability appears to reach 1, judged by eye.

4 Statistical modelling

Turning these historical counts into an exceedance probability curve for the future requires a statistical model. At the global scale, we treat explosive eruptions of magnitude exceeding $M = 4.5$ as an homogeneous (stationary) Poisson process with unknown rate λ (units of yr^{-1}), the rate being effectively constant over the historical time-interval (a, b) plus into the future, where in our case $a = 100$ ky and $b = 2015$ CE. This model has a long history in volcanology; see De la Cruz-Reyna (1991) and the discussion in Rougier et al. (2016).

Our assumption of stationarity deserves some attention as there is strong empirical evidence of local and regional fluctuations of eruption rates, in particular related to glacial and interglacial cycles (e.g., Nowell et al., 2006; Huybers and Langmuir, 2009; Watt et al., 2013; Rawson et al., 2016). At high latitudes enhanced volcanism is associated with warming periods and deglaciation, which can

Table 1: Notation for the recorded eruptions by magnitude.



$m_1, \ldots, m_{k+1}$: breaks for the magnitude bins.

$a_1, \ldots, a_k$: a_i is the earliest date at which the recording probability for $M \geq m_i$ is effectively 1. See Figure 3. $\Delta_i = b - a_i$, the length of time between a_i and 'today' ($b = 2015$ CE).

$n_1, \ldots, n_k$: n_i is the number of recorded eruptions in the bin $m_i \leq M < m_{i+1}$ in the time-interval (a_i, b).

$\mu_1, \ldots, \mu_k$: μ_i is defined in (2).

---

be explained by mantle decompression due to unloading, and changes in the stress state of the lithosphere related to unloading. However, after screening for under-
225 recording, most of the records in this study are within the late Holocene (last 2 ka), except for $M \geq 7.5$ eruptions where we have gone back to 100 ky. Thus our study is within a narrow time-window compared to the above cited studies, which investigated non-stationarity in volcanic rates related to much longer periods. For $M \geq 7.5$ the records are from eruptions at low and intermediate latitudes where
230 the direct effects of glacial unloading are greatly diminished or not apparent. It is possible that the approximately 140 m global change in sea level might influence rates at low latitude, although it is unclear whether this would lead to an increase or decrease in rates. There is no evidence to suggest that rates of extreme magnitude eruptions (i.e. $M \geq 7.5$) are non-stationary since 100 ky.

4.1 ML estimation of rates

The magnitudes of the eruptions are treated as IID 'marks' with an unknown distribution function F_M, for which $F_M(m_1) = 0$. According to the Marking Theorem (Kingman, 1993, sec. 5.2), times and magnitudes together comprise a Poisson process over $(a, b) \times (m_1, \infty)$, with mean function $\lambda \cdot \mathrm{d}F_M$.

Continuing with the notation, let there be k bins for magnitude with breaks $(m_1, m_1, \ldots, m_{k+1})$, and let a_i $(i = 1, \ldots, k)$ be the earliest date at which the recording probability for eruptions with magnitude $M \geq m_i$ is effectively 1; let $\Delta_i = b - a_i$, the length of time between a_i and b. Denote the observations as $(n_1, \ldots, n_k)$, where n_i is the number of recorded eruptions in the set $(a_i, b) \times (m_i, m_{i+1})$. See Table 1 for our notation.

The quantity n_i is Poisson-distributed with expectation equal to $\Delta_i \, \mu_i$, where

$$\mu_i := \lambda \int_{m_i}^{m_{i+1}} \mathrm{d}F_M(m), \tag{2}$$

and, consequently,

$$\lambda = \sum_{i=1}^{k} \mu_i. \tag{3}$$

As these sets are disjoint, the likelihood function for $(\mu_1, \ldots, \mu_k)$ is

$$L(\mu_1, \ldots, \mu_k) \propto \prod_{i=1}^{k} \mathrm{Pois}\left(n_i; \Delta_i \, \mu_i\right), \tag{4}$$

where 'Pois' is the Poisson probability mass function (PMF), with specified expectation. Under this model, the maximum likelihood (ML) estimator for μ_i is

$$\hat{\mu}_i = \frac{n_i}{\Delta_i}, \qquad i = 1, \ldots, k. \tag{5}$$

This estimator will tend to overfit, for example by setting $\hat{\mu}_i = 0$ if $n_i = 0$. However, it is very intuitive, and is much used in practice, possibly without appreciating the statistical model, the Poisson process theory, and the estimation theory

which justify it. More useful is a 95% confidence interval for each μ_i. The problem of choosing a confidence procedure for a Poisson model is still open. We use the procedure originally proposed by Garwood (1936) and adapted by Blaker (2000): see Swift (2009) for a discussion.

Figure 4 shows the ML estimates and confidence intervals for the μ_i's. This figure has two striking features. First, the smoothness of $\log_{10} \hat{\mu}_i$ as a function of m_i: the relationship looks nearly linear. Second, on the basis of that smoothness, additional evidence that the 4th bin, namely $7.5 \leq M < 8.5$, has several unrecorded eruptions since 100 ky, given that the estimate and confidence interval of μ_4 appear to be displaced downwards relative to the smooth relationship of the other estimated μ_i values. There is also a suggestion that μ_3 might be displaced upwards, so possibly some of the eruptions recorded as $6.5 \leq M < 7.5$ should have been recorded as $7.5 \leq M < 8.5$. But only one eruption in the 3rd bin is recorded at $M = 7.4$, Changbaishan (on the border of China and N Korea, dated to 946 CE), and the next-largest is $M = 7.1$.

We cannot think of a physical reason which would lead to a kink around $M = 7.5$, and suggest that there are systematic biases in the estimate of volumes and therefore magnitudes for very large explosive eruptions. All eruptions with $M \geq 7$ form calderas and there are typically three components to the deposits, namely outflow ignimbrites, intracaldera infills and very extensive tephra fall deposits. Johnston et al. (2014) raised the DRE (dense rock equivalent) volume for the Minoan eruption of Santorini from 60 km^3 (Sigurdsson et al., 2006) to 78–86 km^3 with the addition of the intracaldera pyroclastic deposits. This volume change equates to a magnitude change from 7.1 to 7.3. Likewise the proportion of distal tephra fall deposits turns out to be comparable to the proximal ignimbrite volumes for those cases where the deposits have been studied in detail. So a systematic underestimate of volumes for $M \geq 7$ can explain some of the discrepancy, but there are not enough eruptions at the top of the 3rd bin to explain it all.

Another possibility is that there is some non-stationarity at work and that this

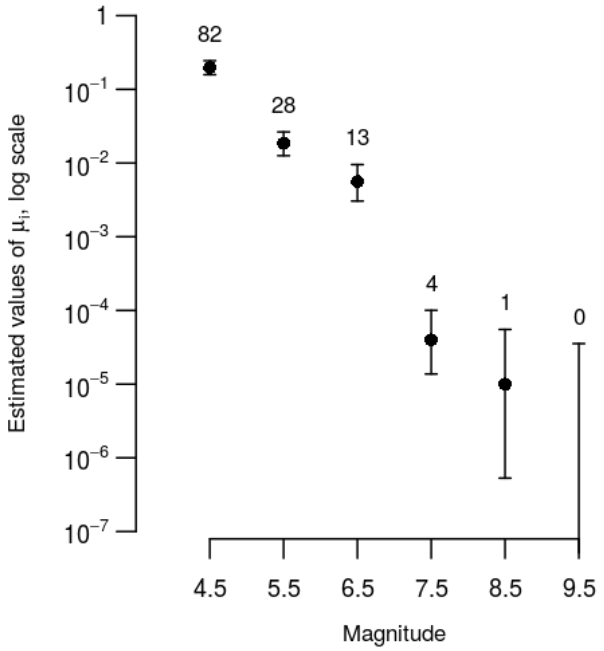


Figure 4: Maximum likelihood (ML) estimates of μ_i computed using (5) (dots), and 95% confidence intervals (error bars). The estimates are plotted against the lower end of their bins (i.e., $\hat{\mu}_i$ is plotted at m_i, see Table 1). The dot at $(9.5, \log_{10}(0))$ cannot be shown. The value above each bar shows the number of records in the bin.

is manifested in the very different time windows we use for the $6.5 \leq M < 7.5$ bin and the $7.5 \leq M < 8.5$ bin. For example if the eruption rates in the last 10 ka have been higher than in the previous 90 ka then the count in the $M < 7.5$ bins would be displaced upwards relative to the $M \geq 7.5$ bins. However, as we explain in section 4, we do not think that non-stationarity is a major issue for our analysis. Therefore, on balance, we favour the idea that there are several $7.5 \leq M < 8.5$ eruptions since 100 ky waiting to be identified.

Ultimately, Figure 4 is slender grounds on which to start moving records between bins. Nevertheless, the numbers of records in these bins are influential in estimating the return period of very large eruptions, and so the possibility of error should not be ignored. Below, we will adopt a weaker hypothesis, that the number of records in $6.5 \leq M < 7.5$ is possibly an over-count, and the number of records in $7.5 \leq M < 8.5$ is possibly an under-count.

Readers wanting simple estimates of exceedance probabilities and return periods should note that the exceedance probability for m_i is approximately equal to μ_i, and the return period for m_i is approximately equal to $1/\mu_i$. These approximations follow from the exact formulae given in section 6. Therefore Figure 4 also provides approximate ML estimates for exceedance probabilities (directly) and return periods (taking the reciprocal). For example, the exceedance probability for $M = 8$ is, by eye, about 5×10^{-5}, and the return period for $M = 8$ is therefore about 20 ka. This suprisingly low value for the return period of $M = 8$ will be confirmed in our more detailed analysis below, and discussed in section 7.

5 Parametric model

The previous assessment, including Figure 4, was semi-parametric, assuming a homogeneous Poisson process for global large explosive eruptions, but making no further assumption about the nature of the magnitude distribution F_M. However, a parametric model for F_M allows us to interpolate the point estimates in Figure 4

to all intermediate values of m, to recognise the possible mis-recording in the two bins $6.5 \leq M < 7.5$ and $7.5 \leq M < 8.5$, and to impose a finite upper limit on M.

In the statistical model, F_M represents the aggregate properties of many volcanoes. An understanding of the physics of a single volcano (see, e.g., Cashman and Sparks, 2013) does not necessarily translate into constraints on F_M. Consider the simple case in which there are two volcanoes, with eruption rates λ_1 and λ_2, and magnitude distribution functions F_1 and F_2. Using the Poisson process model for each volcano, and invoking the Superposition Theorem (Kingman, 1993, sec. 2.2),

$$\begin{aligned} \lambda &= \lambda_1 + \lambda_2, \quad \text{and} \\ F_M &= \frac{\lambda_1}{\lambda} F_1 + \frac{\lambda_2}{\lambda} F_2. \end{aligned} \tag{6}$$

This result generalises immediately to any number of volcanoes. Therefore F_M is a convex combination of the distribution functions of all of the volcanoes, and as such it will tend to be much smoother than the distribution function of any individual volcano. For example, if volcanoes of class A have an interesting kink in their distribution function at magnitude m, then this will be smoothed out in F_M when combined with volcanoes from other classes where there is no kink at m. This result justifies adopting a smooth parametric model for F_M, but at the same time it limits the insight we can derive about an individual volcano, on the basis of the estimated F_M.

We choose the Generalized Pareto distribution (GPD) as our model for F_M, a two-parameter distribution with positive support. First, the GPD has the capability to be linear for low values of m, as suggested by Figure 4, although it would not be linear in general. Second, it has a closed-form expression for its distribution function, which is very convenient for calculations. The GPD model for F_M is chosen for empirical and practical reasons, *not* for its connection with

the theory of extreme values. After truncating at $M = m_u$,

$$F_M(m) = \frac{\mathrm{GPD}(m - m_1; \sigma, \xi)}{\mathrm{GPD}(m_u - m_1; \sigma, \xi)} \tag{7}$$

for $m \leq m_u$, and 1 above, where the GPD distribution function is

$$\mathrm{GPD}(x; \sigma, \xi) := 1 - \left\{1 + \xi \cdot (x/\sigma)\right\}^{-1/\xi} \tag{8}$$

subject to the limits of 0 and 1, where $\sigma > 0$ is a scale parameter and ξ is a shape parameter.

We impose the limit $m_u = 9.3$, which we consider to be a conservative upper bound for maximum explosive eruption size; this is similar to Mason et al. (2004, p. 743), who suggest an upper bound of 9.2. The largest known explosive eruption is Fish Canyon Tuff (27.8 My), with an erupted mass of 1.8×10^{16} kg ($M = 9.2$, Lipman, 1997), which equates to approximately 7.2×10^3 km^3 (assuming a magma density of 2500 kg/m^3). For comparison, current estimates of crustal melt stored within the Yellowstone magmatic system are $< 10^3$ km^3 (Farrell et al., 2014; Huang et al., 2015), and the largest known melt reservoir in the crust—the Altiplano Puna Magma Body in the Andes—may exceed 10^5 km^3 (Ward et al., 2014; Comeau et al., 2015). In both cases, however, the melt resides within a mostly crystalline 'mush' region and is therefore not accessible to a single volcanic eruption (Cashman et al., 2017). Table 3 of Bryan et al. (2010) is a compilation of the largest known silicic eruptive units from large igneous provinces (LIPs), and the very largest of these is recorded at $M = 9.33$ (Paraná-Etendeka, 132 My). As a sensitivity analysis, we also truncated at other values of m_u, but there was no discernable effect on the fitted exceedance probabilities below $M = 8.5$; see Figure 6(a), below.

The two bins $6.5 \leq M < 7.5$ and $7.5 \leq M < 8.5$ have a non-standard treatment in the likelihood function, with the number of records in the latter bin which are wrongly allocated to the former bin treated as uncertain and integrated

out. We assume that this number can be 0, 1, or 2, with equal probability. In addition, the recorded number of eruptions in $7.5 \leq M < 8.5$ plus the number that transfer in from the previous bin is treated as a lower bound on the actual number of eruptions in $7.5 \leq M < 8.5$. In other words, we allow that there might be more missing records in $7.5 \leq M < 8.5$ than just the number that transfer in from the previous bin.

A Frequentist inference is more complicated within this model, and we prefer not to rely on asymptotic approximations. Therefore we switch to a Bayesian inference with the vague prior density function

$$\pi(\lambda, \sigma, \xi) \propto \lambda^{-\frac{1}{2}}/\sigma \tag{9}$$

on the parameter space, and zero outside it. $\lambda^{-\frac{1}{2}}$ is the Jeffreys prior for the Poisson model; $1/\sigma$ is a standard prior for a scale parameter, and ξ has a uniform prior. Choices such as these tend to have credible intervals with accurate coverage properties in the Frequentist sense, as will be confirmed in our application (see Figure 5). For point estimates we use the maximum *a posteriori* (MAP) estimator, while for uncertainties we use 95% equitailed credible intervals (CIs) from the marginal posterior distribution. The resulting values are $\lambda = 0.22$ yr^{-1} (95% CI: 0.18, 0.26), $\sigma = 0.49$ (0.38, 0.59), and $\xi = -0.026$ (-0.089, 0.056). According to these estimates, globally an explosive eruption of $M \geq 4.5$ happens on average about once every five years.

6 Exceedance probabilities and return periods

Let

$$\lambda(m) := \lambda \int_m^\infty \mathrm{d}F_M(m'). \tag{10}$$

Under the Poisson process model in section 4, the exceedance probability and

return period are

$$\bar{P}(m) = 1 - \exp\{-\lambda(m)\} \tag{11a}$$

$$R(m) = \lambda(m)^{-1} \tag{11b}$$

for $m \geq m_1$. When $\lambda(m) \ll 1$, say $\lambda(m) < 0.1$, $\bar{P}(m) \approx \lambda(m)$, and hence $R(m) \approx 1/\bar{P}(m)$. The exact expression in (11a) is used to estimate the exceedance probability curve, given in Figure 5, and the approximation is used to deduce return period estimates, given in Table 2, as discussed in section 2.

The ML estimator of the μ_i's provides an approximate ML estimator of the exceedance probabilities at the m_i values ($i = 1, \ldots, k$), based on

$$\begin{aligned} \bar{P}(m_i) &= 1 - \exp\left\{-\sum\nolimits_{j \geq i} \mu_j\right\} \\ &\approx 1 - \left\{1 - \sum\nolimits_{j \geq i} \mu_j\right\} \\ &= \sum\nolimits_{j \geq i} \mu_j \approx \mu_i, \end{aligned} \tag{12}$$

assuming that $\mu_1 \ll 1$ and $\mu_{i+1} \ll \mu_i$.

Figure 5 shows the estimated exceedance probability curve. One prominent feature is the non-linearity. We show that this is a consequence of the data, and not a necessary feature of the truncated GPD model for F_M. Figure 6(b) shows error bars from a *synthetic dataset* for which the relationship should be linear. The estimated log exceedance probability curve is indeed linear up until the very high values of magnitude where the truncation at $M = 9.3$ forces it to turn downwards.

As an aside, the truncated GPD model might be useful in seismology, where there is thought to be a strongly linear relationship, as embodied by the Gutenberg-Richter law. There, as here, the truncation point must be imposed, but a sensitivity analysis (e.g., Figure 6(a)) can be used to trace its effect back towards magnitudes of more direct concern.

A comparison of the width of the 95% confidence intervals (error bars) and

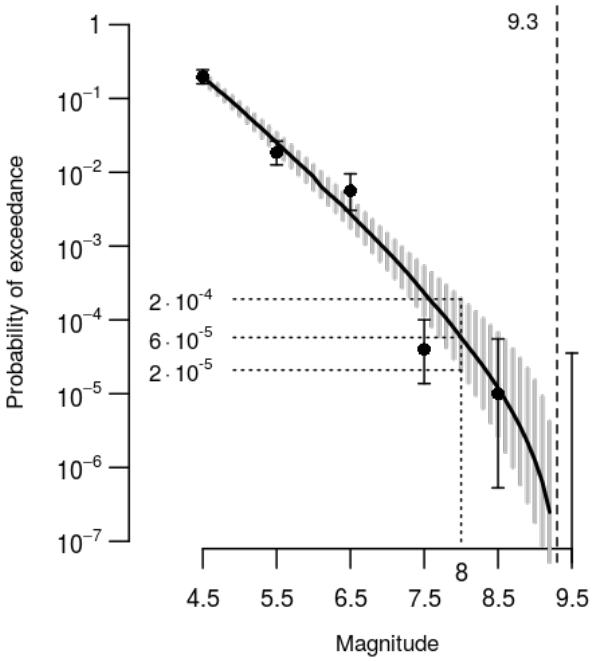
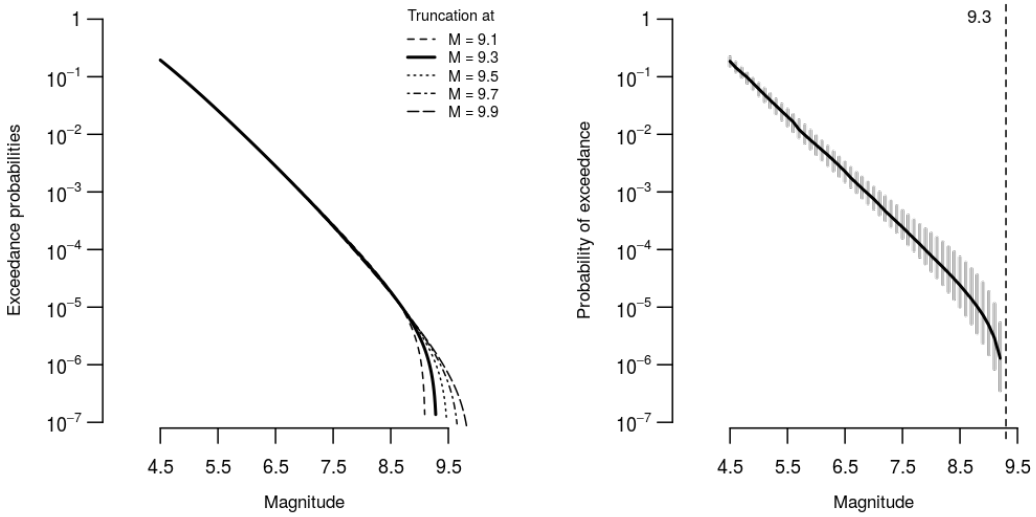


Figure 5: Exceedance probability curve, using a fully parametric approach to F_M (section 5). The solid line is the maximum *a posteriori* (MAP) estimate, and the grey bar is the pointwise 95% credible interval, both based on the vague prior density function given in (9). The error bars are from Figure 4, justified by the approximation in (12). The dotted grid lines indicate the exceedance probability for $M = 8$.



(a) Sensitivity analysis to examine the effect of changing the upper bound on magnitude. The estimated exceedance probabilities below $M = 8.5$ are not affected by the choice of upper bound.

(b) *Synthetic dataset*, to demonstrate the capacity of the truncated GPD model for F_M to fit a linear relationship between magnitude and log exceedance probability. Thus the curved relationship in Figure 5 is a consequence of the data, not the choice of model.

Figure 6: Additional tests for the exceedance probability curve in Figure 5.

95% credible intervals (grey bars) shows that the latter are slightly narrower, notably for large magnitudes. This is expected, because the credible intervals use all records, not just those in a bin. The scarcity of large-magnitude eruptions makes this difference more prominent at large magnitudes.

Figure 5 also highlights the downward displacement of the number of eruptions in the fourth bin ($7.5 \leq M < 8.5$), and the possible upward displacement of the number of eruptions in the third bin. Crudely, it looks as though number of eruptions in the fourth bin should be about three times larger. There are currently 4 eruptions in this bin (Figure 4), suggesting that the number of eruptions with $7.5 \leq M < 8.5$ waiting to be identified is about 8, or fewer if some of the eruptions in the previous bin have been mis-allocated. This seems high, but it is consistent with the size of the gaps in the bottom-righthand frame of Figure 2.

Table 2 contrasts our results with other estimates of the return period by magnitude. There are sizable differences at all magnitudes across the estimates; ours are similar to those of Siebert et al. (2010) for $M \leq 7$, and substantially

different from those of Mason et al. (2004) for $M = 8$.

There are several reasons to expect differences between our estimates and previous values, including that we have used different time periods, and that, compared to Pyle (1995) and Mason et al. (2004), we have used a more modern database.

We suspect that another reason for the $M \leq 7$ divergence is the magnitude-rounding discussed in section 3, and shown in Figure 1. Compared to Sheldrake and Caricchi (2017), who use the same version of LaMEVE as we do (although without our update, see section 3), our return periods for $M \leq 7$ are much longer. Sheldrake and Caricchi (2017) noted the rounding issue, and attributed it, as we do, to rounding to the nearest integer. They also binned their magnitudes, but centred on the 0.5's, not the integers; i.e. they used $4 \leq M < 5$, etc. These are sensible bins for rounding down, but they cause a systematic bias in the presence of rounding to the nearest integer: more records get rounded into the bin $5 \leq M < 6$ from $4.5 \leq M < 5$ than get rounded out of the bin from $5.5 \leq M < 6$, and so on. The estimated exceedance probability curve is pushed upwards, leading to shorter return periods. We identify a source of rounding down and eliminate it from our analysis, through our screening for under-recording (section 3); we treat what remains as rounding to the nearest integer, and use integer-centred bins. But, as we state in section 3.1, we really need a better operational understanding of rounding.

The database compiled by Mason et al. (2004) uses 42 $M \geq 8$ eruptions over the past 36 Ma; that is, they use a much longer time scale than our study. This introduces two problems, as the authors recognise. First, although there is no good estimate of under-recording probabilities over these time scales, the under-recording is likely to be severe. For example, some older ignimbrites can be eroded, buried or incorporated into complex orogenic deformation belts (Wilson, 1991; van Zalinge et al., 2016). Failure to account for under-recording would lead to an artificially long return period. Second, the stationarity of eruption

Table 2: Estimates of the global return period in years for large explosive eruptions.

Magnitude (or VEI)	Pyle (1995)	Mason et al. (2004)	Siebert et al. (2010)	Deligne et al. (2010)	Sheldrake et al (2017)	Us	95% CI
5	8		10	8	6	14	11, 17
6	59		200	35	51	110	80, 170
7	420		1–2 ka	370	420	1200	680, 2100
8		45–714 ka				17 ka	5.2 ka, 48 ka

Pyle (1995): eq. 6, p563, based on VEI and using a density of $2500\,\mathrm{kg/m^3}$.
Mason et al. (2004): p745, λ for $M \geq 8$ is $1.4\text{–}22 \times 10^{-6}\,\mathrm{yr}^{-1}$.
Siebert et al. (2010): p38, based on VEI.
Deligne et al. (2010): Table 6, p14, Holocene, $u = 4.0$; see their Figure 10 for confidence intervals.
Sheldrake and Caricchi (2017): Based on the text above their Figure 4, applying the percentage differences to the values from Pyle (1995).

rates over long time-periods is questionable. For example, the 36 Ma window includes two ignimbrite 'flare-ups' (Lipman, 1984; de Silva and Gosnold, 2007), which suggests that over time-periods of millions of years, the rate of very large-magnitude eruptions may reflect changes in regional tectonics.

By only using $M \geq 8$, Mason et al. (2004) do not constrain their return period estimate with smaller-magnitude eruptions. This constraint is helpful under the assumption of smoothness which we discuss in section 5, but it would not be appropriate under the hypothesis that mechanisms for $M \geq 8$ eruptions are fundamentally different from those for smaller-magnitude eruptions.

7 Conclusion and discussion

We have derived new estimates of the global magnitude-frequency relationship for large explosive volcanic eruptions, presented in terms of the exceedance probability curve, and also summarized in terms of return periods (section 2). Our headline number is that the return period for a $M = 8$ eruption is 17 ka (95% CI: 5.2 ka, 48 ka), much shorter than the previous estimate, but all of our estimates differ substantially from previous estimates.

Our estimates are largely empirical, based on the records in the LaMEVE database, interpreted within an homogeneous Poisson process model. They differ from previous results mainly in our conservative treatment of magnitude-rounding and under-recording (section 3). The semi-empirical results can be seen in Figure 4. This figure already contains the kernel of our estimated exceedance probability curve (Figure 5), and return periods (Table 2). We prefer to use a parametric model for the reasons given at the start of section 5, but these are 'second order' corrections, as is clear from a comparison of the figures. Nevertheless, we believe that these corrections are important, and the results derived from the parametric model are the results we favour, particularly for quantifying variability.

Plag et al. (2015) provide an up-to-date assessment of the risk posed by geohazards, particularly extreme events. They identify volcanoes (and bolides) as hazards capable of producing events large enough to "return humanity to a pre-civilisation state" (Plag et al., 2015, p11). Plag et al. (2015) compute, on the basis of the Mason et al. (2004) return period range for $M = 8$ eruptions ('super-eruptions'), that the expected benefit of a global volcano monitoring system is at least ten times the total cost, and could be "hundreds or thousands of times greater than the total cost" (*ibid.*, p39). On this basis, they assert that humanity is under-prepared for extreme geohazard events (see their Summary of Key Findings, p6, and Conclusions and Recommendations, p9).

Our analysis has produced a much shorter return period for super-eruptions:

a point estimate of 17 ka compared to a lower bound of 45 ka in Mason et al. (2004). So volcanoes are even riskier than previously thought, and Plag et al. (2015)'s assessment that humanity is under-prepared for extreme geohazard events like super-eruptions holds even more strongly. This low value of 17 ka also has important implications for other areas of risk management, which we will explore elsewhere. Briefly, though, we question whether it is cost-effective to manage a risk down to a probability of exceedance of less than $1/(17 \times 10^3)$, if its impact on an entity (such as a country) is much smaller than the impact of a super-eruption happening somewhere in the world.

Therefore our results are interesting not just to volcanologists, but also much more widely, to policy-makers and planners involved in disaster risk reduction (DRR), and to regulators and risk managers.

Acknowledgements

We would like the acknowledge the contribution of the many volcanologists and geologists whose fieldwork has contributed to our modern databases of past eruptions. We would like to thank Natalia Deligne for several very helpful conversations, and two reviewers for their very constructive comments on the first version of this paper.

Funding. Rougier's research was supported by the Natural Environment Research Council (NERC) Innovation Internship scheme (grant number NE/P013155/1). Sparks acknowledges funding from the European Research Council in the VOLDIES project which led to the development of the LaMEVE database, and the Leverhulme Trust (grant number RPG-2015-246) for supporting the management and development of the LaMEVE database. Cashman acknowledges support from the AXA Research Fund.

References

Blaker, H. (2000). Confidence curves and improved exact confidence intervals for discrete distributions. *Canadian Journal of Statistics*, 28:783–798.

Brown, S., Crosweller, H., Sparks, R., Cotterell, E., Deligne, N., Guerrero, N., Hobbs, L., Kiyosugi, K., Loughlin, S., Siebert, L., and Takarada, S. (2014). Characterisation of the Quaternary eruption record: Analysis of the Large Magnitude Explosive Volcanic Eruptions (LaMEVE) database. *Journal of Applied Volcanology*, 3(5). `http://www.appliedvolc.com/content/3/1/5`.

Bryan, S., Peate, I., Peate, D., Self, S., Jerram, D., Mawby, M., Marsh, J., and Miller, J. (2010). The largest volcanic eruptions on earth. *Earth-Science Reviews*, 102:207–229.

Cashman, K. and Sparks, R. (2013). How volcanoes work: A 25 year perspective. *Geological Society of America Bulletin*, 125(5–6):664–690.

Cashman, K., Sparks, R., and Blundy, J. (2017). Vertically extensive and unstable magmatic systems: A unified view of igneous processes. *Science*, 355.

Comeau, M., Unsworth, J., Ticona, F., and Sunagua, M. (2015). Magnetotelluric images of magma distribution beneath Volcan Uturuncu, Bolivia: Implications for magma dynamics. *Geology*, 43(3):243–246.

Costa, A., Smith, V., Macedonio, G., and Matthews, N. (2014). The magnitude and impact of the youngest Toba Tuff super-eruption. *Frontiers in Earth Science*, 2:16.

Crosweller, H., Arora, B., Brown, S., Cotterell, E., Deligne, N., Guerrero, N., Hobbs, L., Kiyosugi, K., Loughlin, S., Lowndes, J., and Nayembil, M. (2012). Global database on Large Magnitude Explosive Volcanic Eruptions (LaMEVE). *Journal of Applied Volcanology*, 1(4). `http://www.appliedvolc.com/content/1/1/4`.

De la Cruz-Reyna, S. (1991). Poisson-distributed patterns of exposive eruptive activity. *Bulletin of Volcanology*, 54(1):57–67.

de Silva, S. and Gosnold, W. (2007). Episodic construction of batholiths: Insights from the spatiotemporal development of an ignimbrite flare-up. *Journal of Volcanology and Geothermal Research*, 167:320–335.

Deligne, N., Coles, S., and Sparks, R. (2010). Recurrence rates of large explosive volcanic eruptions. *Journal of Geophysical Research*, 115:B06203.

Engwell, S., Aspinall, W., and Sparks, R. (2015). An objective method for the production of isopach maps and implications for the estimation of tephra deposit volumes and their uncertainties. *Bulletin of Volcanology*, 77:61.

Farrell, F., Smith, R., Husen, S., and Diehl, T. (2014). Tomography from 26 years of seismicity revealing that the spatial extent of the Yellowstone crustal magma

reservoir extends well beyond the Yellowstone caldera. *Geophysical Research Letters*, 41:3068–3073.

Garwood, F. (1936). Fiducial limits for the Poisson distribution. *Biometrika*, 28(3/4):437–442.

Guttorp, P. and Thompson, M. (1991). Estimating second-order properties of volcanicity from historical data. *Journal of the American Statistical Association*, 86:578–583.

Huang, H.-H., Lin, F.-C., Schmandt, B., Farrell, J., Smith, R., and Tsai, V. (2015). The Yellowstone magmatic system from the mantle plume to the upper crust. *Science*, 348:773–776.

Huybers, P. and Langmuir, C. (2009). Feedback between deglaciation, volcanism, and atmospheric CO2. *Earth and Planetary Science Letters*, 286(3):479–491.

Johnston, E., Sparks, R., Phillips, J., and Carey, S. (2014). Revised estimates for the volume of the late Bronze Age Minoan eruption, Santorini, Greece. *Journal of the Geological Society of London*, 171(4):583–590.

Kingman, J. (1993). *Poisson Processes*. Oxford University Press, Oxford, UK.

Lipman, P. (1984). The roots of ash-flow calderas in western North America: windows into the tops of granitic batholiths. *Journal of Geophysical Research*, 89 (B10):8801–8841.

Lipman, P. (1997). Subsidence of ash-flow calderas: relation to caldera size and magma chamber geometry. *Bulletin of Volconology*, 59(4):198–218.

Mason, B., Pyle, D., and Oppenheimer, C. (2004). The size and frequency of the largest explosive eruptions on earth. *Bulletin of Volcanology*, 66(8):735–748.

Newhall, C. and Self, S. (1982). The Volcanic Explosivity Index (VEI): An estimate of explosive magnitude for historical volcanism. *Journal of Geophysical Research*, 87(C2):1231–1238.

Nowell, D., Jones, M., and Pyle, D. (2006). Episodic Quaternary volcanism in France and Germany. *Journal of Quaternary Science*, 21:645–675.

Plag, H., Blocklebank, S., Brosnan, D., Campus, P., Cloetingh, S., Jules-Plag, S., and Stein, S. (2015). Extreme geohazards: Reducing the disaster risk and increasing resilience. Technical report, European Science Foundation. Available at `https://hazdoc.colorado.edu/handle/10590/4136`.

Pyle, D. (1995). Mass and energy budgets of explosive volcanic eruptions. *Geophysical Research Letters*, 22(5):563–566.

Pyle, D. (2000). Sizes of volcanic eruptions. In Sigurdsson, H., Houghton, B., McNutt, S., Rymer, H., and Stix, J., editors, *Encyclopedia of Volcanoes*. Academic Press, London, UK.

Pyle, D. (2016). Field observations of tephra fallout deposits. In Mackie, S., Cashman, K., Ricketts, H., Rust, A., and Watson, M., editors, *Volcanic Ash: Hazard Observation*, pages 25–37. Elsevier Ltd, Amsterdam, Netherlands.

Rawson, H., Pyle, D., Mather, T., Smith, V., Fontijn, K., Lachowycz, S., and Naranjo, J. (2016). The magmatic and eruptive response of arc volcanoes to deglaciation: Insights from southern Chile. *Geology*, 44(4):251–254.

Rougier, J., Sparks, R., and Cashman, K. (2016). Global recording rates for large eruptions. *Journal of Applied Volcanology*, 5:11.

Sheldrake, T. and Caricchi, L. (2017). Regional variability in the frequency and magnitude of large explosive volcanic eruptions. *Geology*, 45(2):111–114.

Siebert, L., Simkin, T., and Kimberly, P. (2010). *Volcanoes of the World*. University of California Press, Berkeley and Los Angeles CA, USA, 3rd edition.

Sigurdsson, H., Carey, S., Alexandri, M., Vougioukalakis, G., Croff, K., Roman, C., Sakellariou, D., Anagnostou, C., Rousakis, G., Ioakim, C., Goguo, A., Ballas, D., Misaridis, T., and Nomikou, P. (2006). Marine investigations of Greece's Santorini volcanic field. *Eos, Transactions American Geophysical Union*, 87(34):337–342.

Swift, M. (2009). Comparison of confidence intervals for a Poisson mean: Further considerations. *Communications in Statistics: Theory and Methods*, 38(5):748–759.

Tatsumi, Y. and Suzuki-Kamata, K. (2014). Cause and risk of catastrophic eruptions in the Japanese Archipelago. *Proceedings of the Japan Academy, Series B*, 90(9):347–352.

van Zalinge, M., Sparks, R., Evenstar, L., Cooper, F., Aslin, J., and Condon, D. (2016). Using ignimbrites to quantify structural relief growth and understand deformation processes: Implications for the development of the Western Andean Slope, northernmost Chile. *Lithosphere*, 9(1):29–45.

Ward, K., Zandt, G., Beck, S., Christensen, D., and McFarlin, H. (2014). Seismic imaging of the magmatic underpinnings beneath the Altiplano-Puna volcanic complex from the joint inversion of surface wave dispersion and receiver functions. *Earth and Planetary Science Letters*, 404:43–53.

Watt, S., Pyle, D., and Mather, T. (2013). The volcanic response to deglaciation: Evidence from glaciated arcs and a reassessment of global eruption records. *Earth-Science Reviews*, 122:77–102.

Wilson, C. (1991). Ignimbrite morphology and the effects of erosion: a New Zealand case study. *Bulletin of Volcanology*, 53(8):635–644.